



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/811,962	03/30/2004	Akihiro Okumura	32178-201736	2884
26694	7590	10/06/2006	EXAMINER	
VENABLE LLP			VAUTROT, DENNIS L	
P.O. BOX 34385				
WASHINGTON, DC 20043-9998			ART UNIT	PAPER NUMBER
			2167	

DATE MAILED: 10/06/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b> 10/811,962	<b>Applicant(s)</b> OKUMURA ET AL.	
	<b>Examiner</b> Dennis L. Vautrot	<b>Art Unit</b> 2167	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

#### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

- 1) ☒ Responsive to communication(s) filed on 30 March 2004.
- 2a) ☐ This action is **FINAL**.      2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

- 4) ☒ Claim(s) 1-65 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-65 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 30 March 2004 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### Priority under 35 U.S.C. § 119

- 12) ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☒ All    b) ☐ Some \*    c) ☐ None of:
1. ☒ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

#### Attachment(s)

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)  | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)   | 5) <input type="checkbox"/> Notice of Informal Patent Application                       |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)<br>Paper No(s)/Mail Date <u>3/30/2004</u> . | 6) <input type="checkbox"/> Other: _____  |

## **DETAILED ACTION**

### ***Information Disclosure Statement***

1. The information disclosure statement (IDS) submitted on 30 March 2004 has been received and entered into the record. Since the IDS complies with the provisions of MPEP § 609, the references cited therein have been considered by the examiner. See attached forms PTO-1449.

### ***Claim Rejections - 35 USC § 112***

2. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

3. Claims 5, 13, 24, 35, 43, 51, and 59 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

The term "small" in claims 5, 13, 24, 35, 43, 51 and 59 is a relative term which renders the claim indefinite. The term "small" is not defined by the claim, the specification does not provide a standard for ascertaining the requisite degree, and one of ordinary skill in the art would not be reasonably apprised of the scope of the invention. Here, the link depth is rendered indefinite by the use of the term small.

### ***Claim Rejections - 35 USC § 102***

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

5. Claims 1, 2, 7 – 10, 15 and 16 are rejected under 35 U.S.C. 102(b) as being anticipated by **Pirolli et al.** (hereinafter **Pirolli**, US 5,895,470).

6. Regarding claim 1, **Pirolli** discloses an information extracting apparatus for extracting designated information from a document group having a hypertext structure in which documents are mutually related by link information (See column 6, lines 8 – 10 “Referring to FIG. 2, the walker uses the Hypertext Transfer Protocol (HTTP) to request and retrieve a web page, step 201.”), comprising:

a start point address designating unit [walker] which designates an address of the document serving as a start point where said information is extracted (See column 6, lines 4-7 “The site’s topology is ascertained via ‘the walker’, an autonomous agent that, given a starting point, performs an exhaustive breadth-first traversal of pages within the web locality.” The start point addressing unit is defined in the specification in paragraph [0068] as allowing the user to designate the address of a target document to be extracted, which is what is occurring here.); and

an extracting unit which extracts said information from the target document designated by said start point designating unit (See column 6, lines 15 – 19 “The meta-

Art Unit: 2167

information for the page is also extracted and stored, step 204. The meta-information includes at least the following page meta-information: name, title, list of children (pages associated by hyperlinks), file size, and the time the page was last modified.”) and, if said information could not be extracted from said target document, extracts said information from a related document of said target document on the basis of the address of said document. (See column 6, lines 24 – 26 “The list of pages to request and retrieve is then used to obtain the next page, step 206. The process then repeats per step 202 until all of the pages on the list have been retrieved.”)

7. Regarding claim 2, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

8. Regarding claim 7, **Pirolli** additionally teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 “The list of pages to

Art Unit: 2167

request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.)

9. Regarding claim 8, **Pirolli** additionally teaches said upper document [returned page] is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

10. Regarding claim 9, **Pirolli** additionally discloses a category designating unit which designates a category of the information to be extracted (See column 8, lines 55 – 58 “These functional categories might be defined by a user’s specific set of interests, or the categories might be extracted from the collection itself through inductive techniques.”); and

an extracting unit which extracts the information corresponding to said category from the target document designated by said start point address designating unit (See column 6, lines 15 – 19 “The meta-information for the page is also extracted and stored, step 204. The meta-information includes at least the following page meta-information: name, title, list of children (pages associated by hyperlinks), file size, and the time the page was last modified.”) and, if the information corresponding to said category could

Art Unit: 2167

not be extracted from said target document, extracts said information from the related document of said target document on the basis of the address of said document. (See column 6, lines 24 – 26 “The list of pages to request and retrieve is then used to obtain the next page, step 206. The process then repeats per step 202 until all of the pages on the list have been retrieved.”)

11. Regarding claim 10, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

12. Regarding claim 15, **Pirolli** additionally teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 “The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.)

Art Unit: 2167

13. Regarding claim 16, **Pirolli** additionally teaches said upper document [returned page] is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

14. Claims 39 and 47 are rejected under 35 U.S.C. 102(e) as being anticipated by **Murashita** (US 2004/0019499).

15. Regarding claim 39, **Murashita** discloses an information extracting apparatus for extracting designated information from a document group having a hypertext structure in which documents are mutually related by link information (See page 1, paragraph [0008] “The search engine is a system for registering the document on the Internet and its keyword into a server and a searching information by a keyword inputted by the user and is called an agent, an automatic collecting robot, or the like. The search engine scans the document stored in the server on the Internet and forms a document for displaying and a keyword database for searching.”), comprising:

an extracting unit [information collection apparatus] which extracts target information from said document group and, in the case where addition or updating of a document occurs for said document group, executes an extracting process to which



Art Unit: 2167

such addition or updating is reflected each time said addition or updating occurs, and outputs an extraction result including said target information and its document address (See page 9, paragraph [0167] "As mentioned above, in the information collecting apparatus of the invention, the specific site is monitored as an event collecting destination site, if the information in this event collecting destination site has been updated, the keyword to specify the event such as announcement of a new product, incidence of the new virus, or the like is formed from contents of the update, and the information including the keyword is collected from the information collecting destination site by the keyword.");

an extraction result storing unit which stores the extraction result from said extracting unit as extraction result information (see page 9, paragraph [0171] "In step s11, the documents obtained by the information searching unit 26 by using the keyword are stored in the document storing unit.");

a start point address designating unit which designates an address of a document serving as a start point where said designated information is extracted (See page 9, paragraph [0165] "If the user wants to collect information regarding a computer virus by using the information, in step S1, a URL of an antivirus software developing company is preliminarily registered into the event collecting destination site."); and

a searching unit which extracts information from the document of the document address designated by said start point address designating unit and its related document with reference to the extraction result information in said extraction result storing unit (See page 9, paragraph [0166] "...the useful information showing how to

Art Unit: 2167

cope with the new virus as a user of the personal computer is automatically collected by the search of the information collecting destination site by the keyword such as a virus name or the like extracted by the detection of the incidence of the new virus, and it can be shown to the user.”)

16. Regarding claim 47, **Murashita** teaches a category [keyword] designating unit which designates a category of the information to be extracted (See page 9, paragraph [0167] “... if the information in this event collecting destination site has been updated, the keyword to specify the event such as announcement of a new product, incidence of the new virus, or the like is formed from contents of the update, and the information including the keyword is collected from the information collecting destination site by the keyword.”); and

a searching unit which extracts the information belonging to the category designated by said category designating unit. (See page 9, paragraph [0166] “...the useful information showing how to cope with the new virus as a user of the personal computer is automatically collected by the search of the information collecting destination site by the keyword such as a virus name or the like extracted by the detection of the incidence of the new virus, and it can be shown to the user.”)

### ***Claim Rejections - 35 USC § 103***

17. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

Art Unit: 2167

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

18. Claims 3 – 6 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Pirolli** as applied to claim 1 above, and further in view of **Sweet et al.** (hereinafter **Sweet**, US 2002/0073074).

19. Regarding claim 3, **Pirolli** teaches an information extracting apparatus substantially as claimed. **Pirolli** does not explicitly teach a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Sweet** because both are related to operating on linked documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

20. Regarding claim 4, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

21. Regarding claim 5, **Pirolli** additionally discloses an extracting unit which executes the information extracting process in order of the document in which a value of

Art Unit: 2167

the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.)

22. Regarding claim 6, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

23. Claims 11 – 14 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Pirolli** as applied to claim 9 above, and further in view of **Sweet et al.** (hereinafter **Sweet**, US 2002/0073074).

24. Regarding claim 11, **Pirolli** teaches an information extracting apparatus substantially as claimed. **Pirolli** does not explicitly teach a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which,

in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Sweet** because both are related to operating on linked documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

25. Regarding claim 12, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

26. Regarding claim 13, **Pirolli** additionally discloses an extracting unit which executes the information extracting process in order of the document in which a value of the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.)

27. Regarding claim 14, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is

Art Unit: 2167

then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve." Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

28. Claims 17 – 21 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Pirolli** as applied to claim 9 above, and further in view of **Tsuda** (US 7,003,442).

29. Regarding claim 17, **Pirolli** teaches a method substantially as claimed. **Pirolli** fails to explicitly teach a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; an extracting unit which, in the case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document; and a processing unit which outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

However, **Tsuda** teaches a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; an extracting unit which, in the case where only an extraction result of a lower layer in said layer structure



Art Unit: 2167

exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document (See column 15, lines 24 – 29 “Next, the directory file creating unit 43 determines whether or not s2 is empty (at step s76). When s2 is not empty the directory file creating unit extracts a keyword from s2. Next the directory file creating unit determines whether or not the path field of the keyword u is empty.”); and

a processing unit which outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer. (See column 19, lines 34 – 35 “The outputting unit 164 is used to display query messages to the user and processed results.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the extraction method as disclosed in **Tsuda**, the apparatus can effectively search multiple pages and combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have been motivated to include a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; an extracting unit which, in the case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category

Art Unit: 2167

from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document; and a processing unit which outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

30. Regarding claim 18, **Pirolli** teaches a method substantially as claimed. **Pirolli** fails to explicitly teach a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. However **Tsuda** teaches a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the synthesizing rule as disclosed in **Tsuda**, the apparatus can effectively combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have been motivated to include teach a processing unit which has a predetermined

Art Unit: 2167

synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.

31. Regarding claim 19, **Pirolli** teaches a method substantially as claimed. **Pirolli** fails to explicitly teach a processing unit which forms the character string of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure. However, **Tsuda** teaches a processing unit which forms the character string [keyword] of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure. (See column 18, lines 5 – 8 “The processing unit 121 comprises a keyword trimming unit, a keyword relation extracting unit, a directory file creating unit, a searching unit, and a www sever.” And see column 7, lines 50 – 55 “A keyword able contains combinations of [keyword ID (KID), keyword, reading information a set of higher word lds (UP), a set of lower word lds (DOWN), a set of associative word lds (Rel), a set of equivalent keyword lds (Ea), a path, a new word flag (new)].”) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the coupling of the strings as disclosed in **Tsuda**, the apparatus can effectively combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in

the art would have been motivated to include a processing unit which forms the character string of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure.

32. Regarding claim 20, **Pirolli** teaches a method substantially as claimed. **Pirolli** fails to explicitly teach a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. However **Tsuda** teaches a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the synthesizing rule as disclosed in **Tsuda**, the apparatus can effectively combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have been motivated to include teach a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed

Art Unit: 2167

by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.

33. Regarding claim 21, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

34. Claims 22 – 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Pirolli** in view of **Tsuda** as applied to claim 17 above, and further in view of **Sweet et al.** (hereinafter **Sweet**, US 2002/0073074).

35. Regarding claim 22, **Pirolli** and **Tsuda** teach an information extracting apparatus substantially as claimed. **Pirolli** and **Tsuda** do not explicitly teach a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** and **Tsuda** with that of **Sweet** because the references are related to operating on linked documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

36. Regarding claim 23, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address

of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

37. Regarding claim 24, **Pirolli** additionally discloses an extracting unit which executes the information extracting process in order of the document in which a value of the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.)

38. Regarding claim 25, **Pirolli** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the

Art Unit: 2167

pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

39. Regarding claim 26, **Pirolli** additionally teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 "The list of pages to request and retrieve is then used to obtain the next page, step 206." These are examples of link destination documents included in the related document.)

40. Regarding claim 27, **Pirolli** additionally teaches said upper document [returned page] is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See column 6, lines 12 – 15 "The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve." Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

41. Claims 28 – 32 and 37 – 38 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Pirolli** in view of **Tsuda** as applied to claim 17 above, and further in view of **Kunitake et al.** (hereinafter **Kunitake**, US 2002/0073074).



Art Unit: 2167

42. Regarding claim 28, **Pirolli** and **Tsuda** teach an information extracting apparatus substantially as claimed. **Pirolli** and **Tsuda** do not explicitly teach an extracting unit which, in the case where the extraction result is separated into a plurality of character strings of the extraction result of the lower layer and the extraction result of the upper layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, outputs said plurality of character strings as an extraction result of the lower layer and an extraction result of the upper layer.

However, **Kunitake** teaches an extracting unit which, in the case where the extraction result is separated into a plurality of character strings [instruction strings] of the extraction result of the lower layer and the extraction result of the upper layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, outputs said plurality of character strings as an extraction result [document processing description] of the lower layer and an extraction result of the upper layer.

(See page 12, paragraph [0306] "Next, a document processing description synthesizing unit inputs instruction strings separated from plural original documents or templates, merges and sorts the instruction strings, and outputs a document processing description after conversion and synthesis.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** and **Tsuda** with that of **Kunitake** because the references are related to operating on linked documents and by including the

Art Unit: 2167

character strings as disclosed in **Kunitake**, the apparatus can combine information from various layers of the document all in one result. It is for this reason that one of ordinary skill in the art would have been motivated to include an extracting unit which, in the case where the extraction result is separated into a plurality of character strings of the extraction result of the lower layer and the extraction result of the upper layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, outputs said plurality of character strings as an extraction result of the lower layer and an extraction result of the upper layer.

43. Regarding claim 29, the combination of **Pirolli**, **Tsuda**, and **Kunitake** teaches a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See **Tsuda** column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”)

44. Regarding claim 30, The combination of **Pirolli**, **Tsuda**, and **Kunitake** teaches a processing unit which forms the character string [keyword] of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure. (See

Art Unit: 2167

**Tsuda** column 18, lines 5 – 8 “The processing unit 121 comprises a keyword trimming unit, a keyword relation extracting unit, a directory file creating unit, a searching unit, and a www sever.” And see column 7, lines 50 – 55 “A keyword able contains combinations of [keyword ID (KID), keyword, reading information a set of higher word lds (UP), a set of lower word lds (DOWN), a set of associative word lds (Rel), a set of equivalent keyword lds (Ea), a path, a new word flag (new)].”)

45. Regarding claim 31, the combination of **Pirolli**, **Tsuda**, and **Kunitake** teaches a processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See **Tsuda** column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”)

46. Regarding claim 32, the combination of **Pirolli**, **Tsuda**, and **Kunitake** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to

Art Unit: 2167

a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

47. Regarding claim 37, the combination of **Pirolli**, **Tsuda**, and **Kunitake** additionally teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See **Pirolli**, column 6, lines 24-26 “The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.)

48. Regarding claim 38, the combination of **Pirolli**, **Tsuda**, and **Kunitake** additionally teaches said upper document [returned page] is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

49. Claims 33 – 36 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Pirolli** in view of **Tsuda** and in view of **Kunitake** as applied to claim 28 above, and further in view of **Sweet**.

50. Regarding claim 33, **Pirolli, Tsuda, and Kunitake** teach an information extracting apparatus substantially as claimed. **Pirolli, Tsuda, and Kunitake** do not explicitly teach a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli, Tsuda, and Kunitake** with that of **Sweet** because the references are related to operating on linked documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a

Art Unit: 2167

maximum link depth designating unit which designates a maximum link depth; and an extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

51. Regarding claim 34, the combination of **Pirolli, Tsuda, Kunitake, and Sweet** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

52. Regarding claim 35, the combination of **Pirolli, Tsuda, Kunitake, and Sweet** additionally discloses an extracting unit which executes the information extracting process in order of the document in which a value of the link depth is small. (See **Pirolli**, column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.)

53. Regarding claim 36, the combination of **Pirolli, Tsuda, Kunitake, and Sweet** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

54. Claim 40 is rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** as applied to claim 39 above, and further in view of **Pirolli**. **Murashita** teaches an apparatus substantially as claimed. **Murashita** does not explicitly disclose a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and

Art Unit: 2167

retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** with that of **Pirolli** because both are related to information collecting from hypertext documents and by including the internal and external link discrimination as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

55. Claim 41 is rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** as applied to claim 39 above, and further in view of **Sweet**. **Murashita** teaches an apparatus substantially as claimed. **Murashita** does not explicitly disclose a maximum link depth designating unit which designates a maximum link depth; and a searching unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and an searching unit which, in the case where the information could not be extracted from the target document, recursively executes a



Art Unit: 2167

process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** with that of **Sweet** because both are related to operating on web documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and an searching unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

56. Claims 42 – 44 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** in view of **Sweet** as applied to claim 41 above, and further in view of **Pirolli**.

57. Regarding claim 42, **Murashita** and **Sweet** teach an apparatus substantially as claimed. **Murashita** and **Sweet** do not explicitly disclose a searching unit which discriminates an internal link and an external link on the basis of the document address

of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Sweet**, and **Pirolli** because they are related to operating on web documents and by including link discriminating as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

58. Regarding claim 43, **Murashita** and **Sweet** teach an apparatus substantially as claimed. **Murashita** and **Sweet** do not explicitly disclose a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small. However, **Pirolli** teaches a searching unit which executes the

Art Unit: 2167

information extracting process in order of the document in which a value of the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Sweet**, and **Pirolli** because they are related to operating on web documents and by including the link depth order as disclosed in **Pirolli**, the apparatus can be more efficient by searching closer links, which usually contain more relevant information, first. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small.

59. Regarding claim 44, the combination of **Murashita**, **Sweet**, and **Pirolli** additionally discloses a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are

Art Unit: 2167

not in the web locality are not added to the list, thereby discriminating internal and external links.)

60. Claims 45 and 46 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** as applied to claim 39 above, and further in view of **Pirolli**.

61. Regarding claim 45, **Murashita** teaches an apparatus substantially as claimed. **Murashita** does not explicitly disclose said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. However, **Pirolli** teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 "The list of pages to request and retrieve is then used to obtain the next page, step 206." These are examples of link destination documents included in the related document.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** with that of **Pirolli** because both are related to information collecting from hypertext documents and by including the types of documents as disclosed in **Pirolli**, the apparatus can search both upper and lower level documents. It is for this reason that one of ordinary skill in the art would have been motivated to include said related document includes at least one of a link destination document, a link source document, and an upper document of the target document.

Art Unit: 2167

62. Regarding claim 46, the combination of **Murashita** and **Pirolli** teaches said upper document is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

63. Claim 48 is rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** as applied to claim 47 above, and further in view of **Pirolli**. **Murashita** teaches an apparatus substantially as claimed. **Murashita** does not explicitly disclose a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of

Art Unit: 2167

**Murashita** with that of **Pirolli** because both are related to information collecting from hypertext documents and by including the internal and external link discrimination as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

64. Claim 49 is rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** as applied to claim 47 above, and further in view of **Sweet**. **Murashita** teaches an apparatus substantially as claimed. **Murashita** does not explicitly disclose a maximum link depth designating unit which designates a maximum link depth; and a searching unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and an searching unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion.

Art Unit: 2167

This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** with that of **Sweet** because both are related to operating on web documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and an searching unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

65. Claims 50 – 52 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** in view of **Sweet** as applied to claim 49 above, and further in view of **Pirolli**.

66. Regarding claim 50, **Murashita** and **Sweet** teach an apparatus substantially as claimed. **Murashita** and **Sweet** do not explicitly disclose a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches a searching unit which discriminates an internal link and an external link on the basis of the document address

Art Unit: 2167

of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Sweet**, and **Pirolli** because they are related to operating on web documents and by including link discriminating as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

67. Regarding claim 51, **Murashita** and **Sweet** teach an apparatus substantially as claimed. **Murashita** and **Sweet** do not explicitly disclose a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small. However, **Pirolli** teaches a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding



Art Unit: 2167

larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Sweet**, and **Pirolli** because they are related to operating on web documents and by including the link depth order as disclosed in **Pirolli**, the apparatus can be more efficient by searching closer links, which usually contain more relevant information, first. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small.

68. Regarding claim 52, the combination of **Murashita**, **Sweet**, and **Pirolli** additionally discloses a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

Art Unit: 2167

69. Claims 53 and 54 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** as applied to claim 47 above, and further in view of **Pirolli**.

70. Regarding claim 53, **Murashita** teaches an apparatus substantially as claimed. **Murashita** does not explicitly disclose said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. However, **Pirolli** teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 "The list of pages to request and retrieve is then used to obtain the next page, step 206." These are examples of link destination documents included in the related document.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** with that of **Pirolli** because both are related to information collecting from hypertext documents and by including the types of documents as disclosed in **Pirolli**, the apparatus can search both upper and lower level documents. It is for this reason that one of ordinary skill in the art would have been motivated to include said related document includes at least one of a link destination document, a link source document, and an upper document of the target document.

71. Regarding claim 54, the combination of **Murashita** and **Pirolli** teaches said upper document is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper

Art Unit: 2167

directory. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

72. Claim 55 and 63 – 65 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** as applied to claim 47 above, and further in view of **Tsuda**.

73. Regarding claim 55, **Murashita** teaches an apparatus substantially as claimed.

**Murashita** does not explicitly disclose a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; and a searching unit which, in the case where an extraction result of an upper layer is missing only in an extraction result of a lower layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document, and outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

However, **Tsuda** discloses a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; and a searching unit which, in the case where an extraction result of an upper layer is missing only in an

Art Unit: 2167

extraction result of a lower layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document, and outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer. (See column 15, lines 24 – 29 “Next, the directory file creating unit 43 determines whether or not s2 is empty at step s76). When s2 is not empty the directory file creating unit extracts a keyword from s2. Next the directory file creating unit determines whether or not the path field of the keyword u is empty.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** with that of **Tsuda** because they are both related to hypertext document organization and by including the concept of extracting from different layers as disclosed in **Tsuda**, the apparatus the apparatus can effectively search multiple pages and combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have been motivated to include a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; and a searching unit which, in the case where an extraction result of an upper layer is missing only in an extraction result of a lower layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher

Art Unit: 2167

than that of the extraction result of said lower layer from the related document of said target document, and outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

74. Regarding claim 63, the combination of **Murashita** and **Tsuda** teaches a searching unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See **Tsuda** column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”)

75. Regarding claim 64, the combination of **Murashita** and **Tsuda** teaches a searching unit which forms a character string [keyword] of a processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure. (See **Tsuda** column 18, lines 5 – 8 “The processing unit 121 comprises a keyword trimming unit, a keyword relation extracting unit, a directory file creating unit, a searching unit, and a www sever.” And see column 7, lines 50 – 55 “A keyword able contains combinations of [keyword ID (KID), keyword, reading information a set of higher word

Art Unit: 2167

Ids (UP), a set of lower word Ids (DOWN), a set of associative word Ids (Rel), a set of equivalent keyword Ids (Ea), a path, a new word flag (new)].”)

76. Regarding claim 65, the combination of **Murashita** and **Tsuda** teaches a searching unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See **Tsuda** column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”)

77. Claim 56 is rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** in view of **Tsuda** as applied to claim 55 above, and further in view of **Pirolli**. **Murashita** and **Tsuda** teach an apparatus substantially as claimed. **Murashita** and **Tsuda** do not explicitly disclose a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality

Art Unit: 2167

are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Tsuda** and **Pirolli** because they are related to operating on web documents and by including link discriminating as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

78. Claim 57 is rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** in view of **Tsuda** as applied to claim 55 above, and further in view of **Sweet**. **Murashita** and **Tsuda** teach an apparatus substantially as claimed. **Murashita** and **Tsuda** do not explicitly disclose a maximum link depth designating unit which designates a maximum link depth; and a searching unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and an searching unit which, in the case where the

Art Unit: 2167

information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** and **Tsuda** with that of **Sweet** because the references are related to operating on web documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and an searching unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

79. Claims 58 – 60 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** in view of **Tsuda**, in view of **Sweet**, as applied to claim 57 above, and further in view of **Pirolli**.



Art Unit: 2167

80. Regarding claim 58, **Murashita, Tsuda, and Sweet** teach an apparatus substantially as claimed. **Murashita, Tsuda, and Sweet** do not explicitly disclose a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita, Tsuda, and Sweet** with **Pirolli** because they are related to operating on web documents and by including link discriminating as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

Art Unit: 2167

81. Regarding claim 59, **Murashita, Tsuda, and Sweet** teach an apparatus substantially as claimed. **Murashita, Tsuda, and Sweet** do not explicitly disclose a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small. However, **Pirolli** teaches a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita, Tsuda, and Sweet** with **Pirolli** because they are related to operating on web documents and by including the link depth order as disclosed in **Pirolli**, the apparatus can be more efficient by searching closer links, which usually contain more relevant information, first. It is for this reason that one of ordinary skill in the art would have been motivated to include a searching unit which executes the information extracting process in order of the document in which a value of the link depth is small.

82. Regarding claim 60, the combination of **Murashita, Tsuda, Sweet, and Pirolli** additionally discloses a searching unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information

Art Unit: 2167

extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

83. Claims 61 and 62 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita** in view of **Tsuda** as applied to claim 55 above, and further in view of **Pirolli**.

84. Regarding claim 61, **Murashita** and **Tsuda** teach an apparatus substantially as claimed. **Murashita** and **Tsuda** do not explicitly disclose said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. However, **Pirolli** teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 “The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** and **Tsuda** with that of **Pirolli** because they are related to information collecting from hypertext documents and by including the types of documents as disclosed in **Pirolli**, the apparatus can search both upper and lower level

Art Unit: 2167

documents. It is for this reason that one of ordinary skill in the art would have been motivated to include said related document includes at least one of a link destination document, a link source document, and an upper document of the target document.

85. Regarding claim 62, the combination of **Murashita, Tsuda** and **Pirolli** teaches said upper document is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

### ***Conclusion***

86. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

**Ben-Shaul et al.** (6,976,090) teaches excluding external links, depth level concept and input of a URL

**Stern et al.** (US 2002/0052928) teaches discriminating external and internal links as well as the classification/categorization concept


Art Unit: 2167


Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dennis L. Vautrot whose telephone number is 571-272-2184. The examiner can normally be reached on Monday-Friday 8:30-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Cottingham can be reached on 571-272-7079. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Dv  
26 September 2006

  
JOHN COTTINGHAM  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2100

  
27 September 2006